ABSTRACT

The rapid advancement of Natural Language Processing (NLP) and Large Language Models (LLMs) has enabled the automatic generation of human-like text across diverse domains. However, these models often struggle with factual accuracy, frequently producing hallucinations, fabricated or unverifiable information. This is a critical issue, and especially for low-resource languages like Marathi due to the limited availability of large ready-to-use fact-annotated datasets, sophisticated linguistic tools, and established fact-verification systems present in other high-resource languages. Additionally, high-stakes sectors such as governance, education, journalism, and law are particularly vulnerable to inaccuracies, which can result in misinformation, and damage credibility and public trust.

This research aims to develop a Fact-Aware Text Generation Framework specifically for the Marathi language, designed to produce correct, trustworthy, reliable and contextually relevant content. The framework will be built by systematically analysing factual statements from verified Marathi sources, using contextual cues to construct the first comprehensive fact-annotated dataset in Marathi. This dataset will serve as the foundation for training and evaluating the model, ensuring that generated outputs are grounded in verified facts and significantly reducing the risk of misinformation.

The proposed framework aims to ensure and evaluate factual consistency, incorporating mechanisms to detect and reduce hallucinations by verifying generated text against a fact-annotated dataset. Automated metrics will assess accuracy, clarity, and completeness, along with domain-specific evaluations to ensure that it is usable for governance, health care, and education. The framework generated will lead to usable and trustworthy domain-adaptable Artificial Intelligent systems for Marathi, while serving as a scalable model for other lower-resourced languages, and bridging the gap between linguistic fluency and factual accuracy in AI-generated text.

CONTENTS

Abstract……..…….....……....i

List of Figures…….……..……...……..…….…....…..…..iii

1. Introduction……..……..……..…….……..…….……..……1

2. Aim of Study……...…….…….…….…….……..…2

3. Objectives of study……...…….…….…..…….……..…3

4. Study background (Literature review) and Significance of Study…….…4

5. National status …….……..…….…….……..…..……....…….9

6. International status …….……..…….…….……..…..……...…....9

7. Study Hypothesis …….……....……...…….....……10

8. Study Rationale……..……...……..…...…….……11

9. Research Methodology ……...…….…….…...……...…...12

10. Statistical Plan……..……...……..…...……..…..…….15

11. Limitations of the Study……..……...……..…...……..…….16

12. Expected Outcomes of Study……..……..…...…….……17

13. Conclusion……..……...…….....…….…….18

14. Bibliography……..……...……..…….…….……19

LIST OF FIGURES

Figure 1.1 Block Diagram of Fact-Aware Text Generation Framework ……..…..12

Designing a Fact-Aware Text Generation Framework for Marathi Language

1. Introduction:

At present, Marathi is one of the most widely spoken languages in India, with significant cultural, historical, and administrative importance. It is characterized by morphological richness, grammatical complexity, precise vocabulary, and expressive depth. However, despite its prominence, Marathi is considered a low-resource language within the field of Natural Language Processing (NLP). Unlike English and other resource-rich languages, Marathi lacks large-scale fact-annotated datasets, advanced language processing tools, and dedicated frameworks for fact verification. This scarcity of computational resources poses a significant challenge to the creation of reliable, context-sensitive Artificial Intelligence (AI) systems for Marathi. In critical sectors such as governance, education, journalism, law, agriculture, and public policy, errors in machine-generated text can lead to misinformation, damage credibility, and reduce public trust [14]. Therefore, developing fact-aware frameworks for Marathi is essential to address this gap.

Text Generation, or Natural Language Generation (NLG), is generating understandable, contextually appropriate, and relevant human-like text when given structured or unstructured data [15][27]. In the last several years, advances in AI and Machine Learning (ML) as a part of new Large Language Models (LLMs) with transformers have provided substantial improvements in applications such has summarization, translation, dialogue systems, automated reporting, and question answering [11][14][27]. While improved and producing fluent and grammatically correct outputs, most of these models have challenges with factually accurate outputs, and in many cases, they produce hallucinations, fabricated, unverifiable, or contextually inconsistent statements [15][29]. Furthermore, this challenge is worse for low-resource languages such as Marathi, where a lack of high-quality annotated corpora and fact-verification datasets have made it even more difficult to develop text generation models that are factually consistent.

To address this problem, Fact-Aware Text Generation has emerged as a promising solution, ensuring that generated outputs are grounded in verified, trustworthy information. Research

in English, such as FoRAG [11] and FActScore [18], has shown that fact-awareness significantly improves reliability, but there is currently no dedicated framework for Marathi. This proposed research aims to address this gap by developing a Fact-Aware Text Generation Framework for the Marathi language that ensures factual consistency, contextual accuracy, and credibility. The research involves examining naturally occurring Marathi data to ascertain factual statements [33][34], building the first comprehensive fact-annotated Marathi dataset, and integrating fact checking whenever information is generated, whether it's in the generative or evaluative process.

The proposed framework systematically develops and thoroughly assesses methods to mitigate hallucinations and validate each output with credible sources [15][27]. It combines both linguistic and contextual analysis and annotates facts from a dataset to provide answers that are accurate, reliable, and domain relevant. The fact-aware text generation system will be scalable, resilient and could be applicable in the areas of government, newspapers, education, agriculture and health care. In developing the fact-aware text generation framework specifically for Marathi, the study contributes towards closing the gap between linguistic fluency and fact-checking in AI text generation and laying the groundwork to expand to similar applications to other low-resource languages [33].

2. Aim of the Study:

This research mainly aims to build a fact-aware systematic text generation framework, of Marathi language that guarantees factual accuracy, verifiability, and contextuality. As AI-generated content is becoming prevalent in high-stakes fields such as governmental policy, education policy, media, law, financial services, and public policy, it's imperative for generated text be both fluent, coherent, and a potential claim and grounded in verified facts. The goal of this research is to create a comprehensive fact-annotated dataset from diverse Marathi language sources, including both structured and unstructured data. This dataset will form the foundation for developing a fact-aware text generation framework capable of automatically retrieving, selecting, and integrating verified facts into the generation process. The framework will be designed to produce content that is factually accurate, semantically

relevant, and logically consistent, while aligning with the linguistic characteristics of Marathi.

Further, this study trains and evaluates the proposed framework using the dataset, validating factual consistency through hallucination detection and mitigation methods applied during text generation. The approach also automates evaluation using measures of both factuality and semantic relevance, complemented by human-centered assessments to ensure credibility, accuracy, and domain-specific efficacy. It addresses a critical gap in fact-aware text generation for low-resource, morphologically rich languages and offers a scalable template for similar work in other regional languages.

3. Objectives of study:

The objective of this research is to design a fact-aware text generation framework for the Marathi that creates trustworthy, verifiable, and fact-based content in Marathi language.

The Objective of this Research:

- To explore and analyse facts from textual data through contextual analysis.
- To design and construct a fact-annotated dataset in Marathi for fact-aware text generation and evaluation.
- To train and evaluate a fact-aware framework to generate Marathi text that ensures factual consistency.
- To detect and minimize hallucinations in Marathi text generation by validating content against the fact-annotated dataset.

4. Study background (Literature review) and Significance of Study:

Literature review:

Large Language Models can produce coherent and contextually relevant text, but they often generate information that is factually incorrect when they have no access to quality data. This tendency is particularly problematic for low resource languages like Marathi, which face hurdles in developing an annotated dataset, establishing a fact-verification infrastructure and developing a domain specific generation system. Linyao Yang et al. [1] proposed a knowledge graph integrated generation framework, where LLMs dynamically retrieve facts from structured knowledge bases to reduce hallucination rates. The approach improved factual grounding in English summarization tasks but depends heavily on well-developed language-specific knowledge graphs, which do not exist for Marathi.

Gilad Fuchs et al. [2] developed a practical, fact-checking system for LLMs that allows real-time verification as content is created. They provided a retrieval mechanism combined with rule-based verifications to eliminate false or unverifiable data which is very important when working in mission-critical systems like legal or medical. Lei Huang et al. [3] systematically examined hallucination in LLM, categorizing it into intrinsic and extrinsic. Their taxonomy gave a basis for developing directly targeted strategies for mitigating hallucinations. Rishabh Upadhyay et al. [4] examined retrieval (augmented generation) concerning topical relevance in health information systems and found that retrieval accuracy improved factual correctness significantly. However, applying this work in Marathi context is limited by the availability of multilingual or Marathi-specific retrieval datasets.

Hanikova et al. [5] integrated entity graphs with argumentation frameworks to generate summaries that were both factually accurate and logically coherent. Their method links entity nodes to argumentative claims, enabling better conflict resolution in generated content. Junnan Liu et al. [6] developed S-Evaluator, an adversarial testing platform that challenges LLMs with complex and ambiguous queries to measure factual robustness. By synthesizing adversarial test cases, the tool evaluates a model's ability to resist generating hallucinated content. Muneeswaran et al. [7] proposed mitigation techniques such as targeted prompt engineering and hybrid verification layers to address factual inconsistencies in LLM outputs, with potential adaptation for low-resource languages like Marathi. Haojie Zhuang

et al. [8] improved faithfulness and factuality in recommendation systems using contrastive learning, highlighting that similar approaches could be applied to content generation frameworks. Zheheng Luo et al. [9] presented methods for evaluating factual consistency in summarization, emphasizing precision and recall metrics specifically designed for the era of LLMs. Yunqi Xu et al. [10] introduced Face4RAG, a factual consistency evaluation system tailored for Chinese retrieval-augmented generation, showcasing cross-lingual potential if adapted to Marathi.

Tianchi Cai et al. [11] developed FoRAG, a factuality-optimized RAG model for web-enhanced long-form question answering, which dynamically integrates retrieved information into the generation process. Md Moinul Islam et al. [12] evaluated summarization techniques and their factual consistency when using LLMs, noting the challenges of balancing conciseness with factual completeness. Arghya Kundu et al. [13] applied knowledge-graph-based models for automated fact-checking, demonstrating the value of semantic networks for verifying generated claims. Jiaju Kang et al. [14] demonstrated that entity-augmented correction modules can align LLM outputs with gold-standard facts in domain-specific tasks such as consumer complaints. Despite these advances, no equivalent entity-level hallucination detection or correction system exists for Marathi text generation, leaving a gap in domain-sensitive applications such as governance or education.

Xinyu Li et al. [15] proposed a hallucination detection framework based on fact and semantic consistency, allowing for real-time detection of fabricated content during generation. Hongjin Qian et al. [16] introduced dynamic contextual fact retrieval, which reduces factual drift in long-form outputs by continuously injecting relevant external facts during generation. While effective in English, the lack of large-scale, fact-annotated Marathi corpora prevents adaptation of these strategies. Craig Thomson et al. [17] emphasized aggregate factual integrity in multi-document data-to-text generation, demonstrating that factual alignment across multiple inputs improves decision-making reliability. Despite these contributions, the Marathi language lacks domain-agnostic, multi-source fact-aware frameworks capable of integrating structured and unstructured sources.

Sewon Min et al. [18] introduced FactScore, which evaluates factual correctness at the atomic claim level, outperforming conventional overlap-based metrics. However, these tools

are primarily designed for English, and fine-grained factuality metrics for Marathi are still emerging, making factual evaluation challenging. Derek Tam et al. [19] measured factual consistency in news summarization, revealing that fine-tuned LLMs can maintain higher accuracy with proper contextual grounding. Congqing He et al. [20] applied multi-task learning to enhance factual consistency in legal judgment explanations, combining textual generation with legal reasoning validation.

Yiyang Li et al. [21] surveyed factual consistency methods from 2021 to 2023, categorizing techniques into pre-generation, in-generation, and post-generation strategies. Alejandro Martin et al. [22] developed FacTeR-Check, a semi-automated fact-checking approach that merges semantic similarity analysis with natural language inference. Nayeon Lee et al. [23] designed factuality-enhanced language models for open-ended text generation, embedding verification into the generation pipeline itself. Wenya Li et al. [24] improved factual consistency in dialogue summarization through fact-augmentation mechanisms that incorporate structured knowledge sources. Amanuel Alambo et al. [25] employed multi-objective optimization to balance factual accuracy with narrative readability in clinical summaries. Jeongwan Shin et al. [26] proposed a token-level post-generation rectification method using reinforcement learning to detect and correct erroneous tokens in abstractive summarization tasks. Ruoyu Chen et al. [27] used dependency graphs in generative summarization, enabling fact-aware structuring of generated content.

Ben Goodrich et al. [28] examined methods for assessing factual accuracy of generated text, showing that both automated metrics and human evaluation are necessary for reliable assessment. Yingqi Zhu et al. [29] improved factual error correction using data distillation and cloze-style generation to refine model outputs. Anup Bera et al. [30] validated numerical data in financial summaries, an important area for preventing subtle but impactful factual errors. Huan Rong et al. [31] developed FuFaction, a fuzzy factual inconsistency correction framework that operates at the hidden-state level within neural networks. Yaping Chai et al. [32] reviewed text data augmentation strategies for LLMs, emphasizing their role in reducing hallucination through diversified training data.

Vidula Magdum et al. [33] provides resources for Marathi NLP (tokenization, tagging, embeddings) and baseline models for key NLP tasks, lowering the entry barrier for research in the Marathi context. Raviraj Joshi et al. [34] made available a large Marathi monolingual

corpus and pretrained Marathi BERT models, which substantially improve downstream tasks. These resources are essential for Marathi NLP research. Qianren Mao et al. [35] explored multi-relational knowledge-based summarization to improve factual correctness in generated outputs. V. P. Kadam et al. [36] developed a text summarization system for Marathi, showing the feasibility of adapting summarization pipelines to low-resource languages. Dhiraj Amin et al. [37] studied Marathi–English code-mixed text generation, addressing challenges in multilingual and mixed-lingual contexts common in Indian digital communication.

- **Research Gaps from Literature:**

- Lack of Fact-Aware Frameworks for Marathi Language

While fact-aware text generation frameworks exist for English and other high-resource languages, there is no comprehensive framework addressing factual consistency for Marathi. Existing works, such as FoRAG [11] fact-semantic consistency methods [15] and factuality-enhanced language models [23] are primarily developed for English datasets and tools, making them unsuitable for regional languages with unique linguistic structures and morphological richness.

- Absence of a Fact-Annotated Marathi Dataset

Most fact-checking and factual consistency research relies on datasets in English e.g. FactCC, FEVER. Studies like news summarization factuality checks [19] and fact-aware summarization using dependency graphs [27] demonstrate progress in English, but there is no publicly available fact-annotated corpus for Marathi text generation. This gap limits the ability to train and evaluate Marathi-specific fact-aware models, despite the existence of general Marathi corpora like MahaCorpus [34].

- Limited Hallucination Detection and Mitigation in Low-Resource Languages

Hallucination correction methods, such as entity-augmented hallucination correction [14] fact semantic consistency frameworks [15] and factual error correction via data distillation [29] have been effective for English but are not adapted for Marathi. Hallucinations in generated content can cause misinformation, especially in sensitive domains such as governance, journalism, and education.

- Lack of Post-Generation Fact Validation Pipelines

Most frameworks focus on fact-aware training but lack robust verification mechanisms that validate generated content after generation. Dynamic knowledge selection approaches [16] and factuality evaluation studies [28] show that post-generation checks are essential, yet such pipelines are missing for Marathi.

- Limited Explainability and Source Traceability in Regional Language Text Generation

Some fact-aware models, such as factuality-enhanced language models [23] fact-aware summarization with dependency graphs [27] and multi-relational knowledge-based summarization [35] improve factual accuracy. However, they often lack explainable mechanisms that allow users to trace generated statements back to their original verified sources, a crucial requirement for high-stakes domains such as journalism, governance, and education in Marathi.

• Significance of the Study:

As Large Language Models become more widely applied across many fields, particularly in important areas such as governance, education, journalism, and public communication, the truthfulness, leading to accuracy and reliability and factual grounding in the AI-generated material that we use, is as important a concern as it has ever been [1]. In important fields, a very small amount of inaccuracy is often enough to support misinformation or misinterpretation and may result in a loss of trust in public oversight activity [2]. This issue is likely to be particularly important for low resource and morphologically rich languages like Marathi. To address this, we propose designing fact-aware text generation framework for Marathi language that aims to achieve factual accuracy, contextual relevance, and verifiability of the output from AI-generated output texts [4].

A primary feature of this framework includes a sensitive hallucination detection and suppression process, where generated text is validated against a fact-annotated Marathi resource dataset. This not only increases factual correctness but ensures contextual traceability, which allows users to verify claims to the original sources, an extremely relevant feature for sensitive decision-making situations. [10].

In addition, it is a domain-adaptable framework, which means that it can be applied to a wide range of domains with little to no readjustment [11]. Since none of the studies have addressed fact-aware NLP for Marathi, this work contributes to an important gap in the literature regarding factual AI in low-resource languages and also has provided researchers with a framework for deploying safe and reliable generative AI applications [33].

5. National Status:

India has made considerable progress in AI through strategic initiatives such as AI4Bharat and the National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS), supported by the Department of Science and Technology (DST) and the Ministry of Electronics and Information Technology (MeitY). Leading institutions including IITs, IIITs, CDAC, and IIIT-Hyderabad have advanced NLP by developing multilingual models and tools for India's diverse linguistic ecosystem. While these achievements have strengthened language technologies for Indian languages, most progress has been in translation, summarization, and sentiment analysis, with limited focus on factual accuracy in generated text.

The proposed research seeks to fill this gap, by developing a Fact-Aware Text Generation Framework for Marathi Text Generation, with the aim of producing contextually relevant text outputs which are factually consistent and verifiable. This supports India's aim of being an ethical AI-figuring out and bolstering the reliability of the AI powered language technologies that will be available to the public, and in specified domains.

6. International Status:

Major AI organizations across the world, including OpenAI, Google DeepMind, Meta AI, and Microsoft Research are actively working on improving the factual accuracy of AI generated content. This includes research and development of retrieval-augmented generation models that work to reduce hallucinations of facts by using facts from relevant and verifiable sources of knowledge [5]. Academic institutions including MIT, Stanford University, and Carnegie Mellon University are also contributing to the development of interpretable and verifiable AI systems, reinforcing the global emphasis on ethical,

transparent, and fact-based AI. There are English datasets and evaluation frameworks, such as FEVER, QAGS, and FactCC [17][18] which can measure the degree of factual accuracy in generated text, and is based on upon evidence and multiple sources, and top AI conferences like EMNLP, ACL, and NeurIPS are introducing all new things in fact-aware generation, fact-checking, and knowledge integration.

The research proposal is addressing these global trends by designing a Fact-Aware Text Generation Framework for Marathi language that has factually consistent, contextually relevant and verifiable outputs. With the combination of fact-annotated datasets and the incorporation fact validation, the framework is bringing low-resource, morphologically rich languages like Marathi into international AI discussions, with goals for both technology, as well as ethical standards in artificial intelligence.

7. Study Hypotheses:

In order to verify the feasibility of the proposed fact-aware text generation framework for Marathi and to assess whether it actually provides measurable value over existing text generation approaches, the hypotheses below were developed to provide an objective measure of the empirical findings, rather than just a hypothesis alone.

H_0 (Null Hypothesis):

The proposed framework for fact-aware text generation in Marathi language does not provide any statistically significant enhancement in factual accuracy, contextual relevance, hallucination reduction, and overall trustworthiness compared to existing Marathi text generation systems.

H_1 (Alternative Hypothesis):

The suggested fact-aware text generation framework for Marathi language has produced significant, clear increase in factual accuracy, contextual relevance, hallucination and reliability compared to other Marathi text generation developments.

The stated hypothesis test will assess whether the proposed framework yields measurable advantages over or in real-inspired situations. The tests will be controlled trials in which the proposed framework will be tested, as will existing Marathi text generation models in similar fashion with the same Marathi datasets in each case. Measures of performance will focus on factual consistency scores, contextual fit, hallucination rate, and verifiability of outputs.
Statistical testing methods (e.g., paired t-test, Wilcoxon signed-rank test, or ANOVA) will be applied to compare results. If the improvements achieved by the proposed framework are found to be statistically significant, the null hypothesis (H_0) will be rejected in favour of the alternative hypothesis (H_1), thereby confirming the framework's effectiveness and alignment with the stated research objectives.

8. Study Rationale:

The reason for this study is to address the persistent and critical problem of factual inaccuracy in AI-generated text, with a specific focus on the low-resource and morphologically rich Marathi language. In high-stakes areas such as governance, education, journalism, and policy communications, inaccurate, unverifiable, or fake outputs from LLMs generate misinformation, misinterpretation, and huge societal harm. The aspects of text generation systems tend to view fluency and coherence over factually accurate outputs, and view generation and verification as separate functions in the text processing chain. These aspects based on the limited transparency LLMs provide seem to be why the issue of accountability is so concerning, as the outputs are difficult to trace, validate and explain, which has impacts on trust, accountability, and the responsible development and use of reliable and trustworthy AI technology.
This research aims to fill this gap through the design of a unified fact-aware text generation framework for Marathi, that incorporates factual verification into the generation pipeline. This will include building a fact-annotated dataset for Marathi, using linguistic and contextual features to systematically identify factual statements from Marathi data, and training a model to generate verifiable text that is accurate and contextually relevant. It will pull information from data sources, ensuring that the generated outputs are grounded in reliable and relevant information

A core aspect of this strategy is hallucination detection and minimization, where generated outputs will be assessed against a fact-annotated dataset using both automated verification and human-in-the-loop assessment. This integrated design will improve factual accuracy and contextual relevance while enabling scalable, domain-adaptable fact-aware text generation for other low-resource languages. The results of this study contribute to national and international priorities for ethical, transparent, and trustworthy AI, advancing digital empowerment, preventing misinformation, and fostering the development of AI solutions that are linguistically and culturally equitable.

9. Research Methodology:

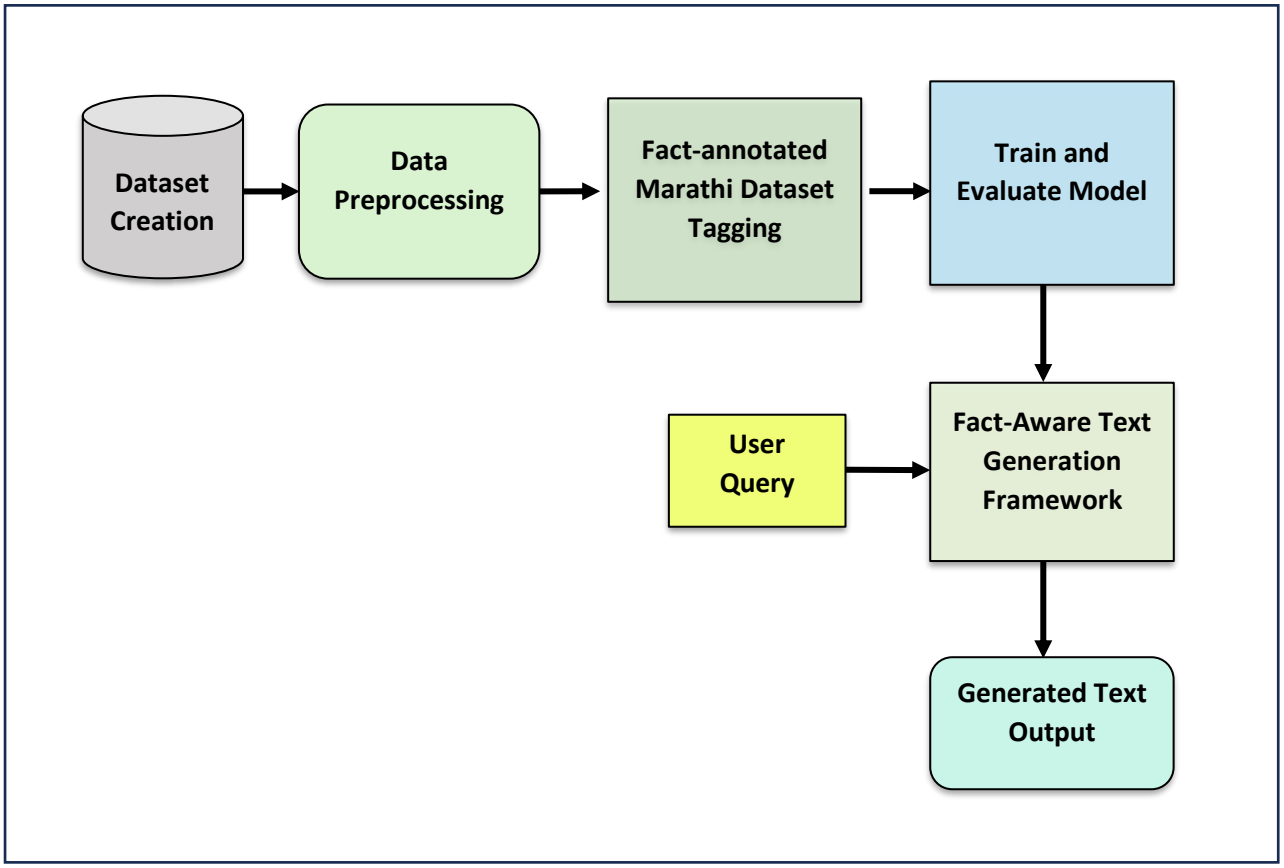

Figure 1.1: Block Diagram of Fact-Aware Text Generation Framework

The proposed Fact-Aware Text Generation Framework (Figure 1.1) for the Marathi Language follows a structured and sequential process designed to produce text that is both factually reliable and linguistically fluent.

The initial stage of proposed work is Dataset Creation, where a large, diverse, and domain-rich Marathi text dataset is created and compiled from trusted and verifiable sources. These sources include data compiled from government publications, reputable news agencies, peer-reviewed research articles, historical archives, and educational repositories to ensure the data is credible and genuine. This way, the dataset not only encapsulates the linguistic richness, morphological richness, and cultural intricacies of Marathi, but achieves factually reliability through source credentials. The collection of sources was also themed with a wide variety governance, healthcare, agriculture, education, law and technology thereby allowing to meet a wide variety of factual queries.

After collection, the data enters the Data Pre-processing phase, wherein raw text is cleaned to remove unwanted noise such as tags, advertisements, incomplete sentences, irrelevant information, and duplicate content. Linguistic normalization is applied to eradicate grammatical and morphological differences, enabling consistency across different dialectical forms and orthographic variations in Marathi. This includes stemming, lemmatization, and grammatical inflections, signaling aspects of Marathi grammar. The tokenization, normalization, and context marking guarantee that auxiliary meaning and semantic dependencies will prevail across the dataset. At this phase, we have achieved the first row of procedures to change the unstructured corpus into a structured, uniform dataset that has been further prepared for fact-tagging and model training.

The next phase is Fact-Annotated Marathi Dataset Tagging, where factual statements are identified, cross-verified with authoritative references, and labeled with fact-tags while being linked to precise source citations. Cross-verification includes automated fact- checking pipelines as well as human validation by expert annotators to ensure factual precision. Fact-tags also include metadata such as temporal validity, domain-specific terminology, and reliability scores to improve accuracy in downstream applications. In the Model Training and Evaluation stage, the fact-annotated dataset is used to train the model, ensuring it learns to generate coherent, contextually relevant, and factually consistent Marathi text. This

training process leverages transformer-based architectures fine-tuned on the fact-annotated corpus.

Once a User Query is submitted, it gets processed inside the Fact-Aware Text Generation Framework that dynamically retrieves relevant and validated information from the fact-annotated dataset. The system filters and matches the information intelligently to assure that the retrieved facts align nicely with the intended goal of the user and no ambiguity exists regarding contextual appropriateness of the query. During the fact's generation, statement is automatically compared with fact-annotated dataset and cross-verified with fact-verification modules. This stage of verification ensures that there is no entry into the output of misleading, outdated, or fabricated information. By implementing this strict validation process, the framework will successfully control and eliminate hallucinations, factual inconsistencies, or unverifiable claims to assure the integrity, accuracy, consistency and trustworthiness to the generated Marathi text.

Finally, the framework generates the Text Output, which is fact-based, contextually relevant, true, trustworthy, and linguistically rich. These outputs are most applicable in high-stakes areas like political governance, journalism, medical healthcare, criminal justice, and education where, credibility, accuracy, and public trust are dominant. By integrating fact verification at multiple stages of the pipeline, this framework not only addresses the limitations of existing Marathi NLP systems but also establishes a scalable, ethical, and reliable AI approach. Moreover, the framework's architecture is designed to be adaptable to other low-resource languages, thus offering a blueprint for broader multilingual fact-aware Artificial Intelligent development.

10. Statistical plan:

The proposed framework for Fact-Aware Text Generation in Marathi language is assessed by both experimental and inferential statistical methods. This provides assurance that the evaluation aligns with the core purposes of the research. Measurement includes factual accuracy, contextual relevance, hallucination reduction, source traceability, scalability, and scrutiny against baseline Marathi text generation controllers as specified in the second research question.

Factual consistency of generated outputs is measured using factuality metrics to examine the correctness of Marathi materials generated through comparison with the fact-annotated dataset. Semantic relevance is measured in terms of text fluency and overall contextual relevance. This includes BERTScore along with ROUGE-L, which evaluate whether the generated outputs maintain sufficient language quality for Marathi text generation, plus general contextual relevance. Outputs are aggregated across representative domains such as governance, healthcare, and education, with measures averaged across the test sets. Source traceability is assessed using precision and recall metrics to validate the correct attribution of factual statements to source documents.

Scalability is analyzed by recording computational time, memory usage, and latency while processing varied input sizes, with results represented using trend plots. Reliability is ensured through human evaluations, in which three expert annotators per domain rate outputs on a 5-point Likert scale for factual accuracy, coherence, and contextual appropriateness; inter-rater agreement is measured using Fleiss' Kappa. Robustness testing introduces controlled perturbations such as partial or conflicting input data—and measures performance degradation via mean squared semantic drift, calculated through cosine similarity in BERT embedding space with respect to reference outputs.

For statistical analysis, one-way ANOVA examines significant differences in factual consistency and semantic relevance across models, while independent t-tests compare the proposed framework with baseline systems at a significance level of $\alpha = 0.05$. Additionally, Spearman correlation assesses the alignment between human judgments and automated metric scores. This rigorous evaluation framework ensures that the system meets its

objectives of delivering factually consistent, contextually accurate, and trustworthy Marathi text generation while effectively minimizing hallucinations.

11. Limitations of the Study:

While the proposed Fact-Aware Text Generation Framework for Marathi aims to address key challenges in factual accuracy, contextual relevance, and hallucination reduction, certain limitations are anticipated. The research may face different limitations, including:

- Language-Specific Resource Constraints – Marathi, being a low-resource and morphologically rich language, lacks large-scale, high-quality, fact-annotated datasets. The creation of such datasets will be time-consuming and may not cover all domains comprehensively, potentially affecting model generalization.
- Domain-Specific Bias – While the framework will aim to be domain-independent, the initial dataset collection may be skewed towards certain sectors (e.g., governance, healthcare, education) where factual sources are more readily available, potentially limiting performance in underrepresented domains.
- Scalability and Computational Requirements – Processing large volumes of multi-source Marathi content, particularly when integrating both structured and unstructured sources, may require high computational resources. This could limit scalability in resource-constrained environments.
- Human Evaluation Constraints – Human-in-the-loop validation is critical for ensuring factual correctness, but reliance on expert annotators may introduce subjective bias and is resource-intensive in terms of time and cost.
- Limited Availability of Automated Fact-Checking Tools for Marathi – Most existing factual consistency evaluation tools are developed for English and other high-resource languages. Their adaptation to Marathi may require significant customization, and some semantic nuances might still be missed.
- Dynamic Nature of Facts – Factual correctness is time-sensitive, especially in fast-changing domains such as news, policies, and regulations. Without regular updates to the fact-annotated dataset, generated content might become outdated or partially inaccurate.

- Transferability to Other Languages – Although the framework could be adapted for other low-resource languages, linguistic differences, cultural contexts, and resource availability might limit its direct applicability without significant retraining and reannotation.

12. Expected Outcomes of the Study:

While this research aims to design a robust fact-aware text generation framework for the Marathi language, certain limitations may influence its scope and applicability:

- Limited Availability of High-Quality Marathi Data – As Marathi is a low-resource language, obtaining large-scale, domain-diverse, and verified factual datasets is challenging. This may impact the depth of training and evaluation compared to high-resource languages.
- Complexity of Fact-Verification in Marathi – Linguistic nuances, morphological richness, and context-dependent meanings in Marathi may complicate the automatic detection and validation of factual statements, leading to occasional misclassification or overlooked factual inconsistencies.
- Domain Coverage Restrictions – Although the framework will be designed to be domain-independent, initial dataset creation and evaluation will be limited to selected domains, which may affect generalization to other domains.
- Dependency on Annotation Quality – The accuracy of the fact-aware framework will heavily rely on the precision and consistency of human annotations in the fact-annotated dataset. Errors or inconsistencies in annotation could propagate into the model's training and evaluation.
- Computational and Resource Constraints – Training and fine-tuning large language models for Marathi, especially with fact-verification modules, require significant computational resources, which may limit experimentation with very large datasets or complex architectures.
- Potential Challenges in Hallucination Detection – While the framework will integrate hallucination detection mechanisms, subtle factual distortions or omissions may still go undetected, especially in complex, multi-source inputs.

- Scalability to Other Low-Resource Languages – Although the framework can serve as a base for other Indian languages, adaptations may require additional linguistic resources, making scalability dependent on language-specific resources and expertise.

13. Conclusion:

The proposed research presents a structured and dependable approach to fact-aware text generation for the Marathi language, aimed at ensuring higher levels of factual consistency, accuracy, and transparency. By leveraging a dedicated fact verification process supported by a fact-annotated Marathi dataset, the framework grounds generated content in verified and contextually relevant information, thereby significantly reducing hallucinations and factual inaccuracies. In addition, the inclusion of source attribution, evidence tracing, and rationale visualization enables users to track the origin of each generated statement and understand the reasoning behind it, an essential capability for high stakes domains such as healthcare, governance, legal systems, and education, where credibility and interpretability are critical.

Beyond improving factual accuracy, this study contributes to the development of a scalable, adaptable, and domain-independent framework that can generate verifiable and context-rich Marathi text. By combining linguistic and contextual analysis, a dedicated fact-annotated dataset, and rigorous evaluation mechanisms, the research directly addresses the risk of misinformation and supports the creation of reliable, ethical, and accountable AI systems. While primarily focused on Marathi, the proposed framework has the potential to serve as a reference model for other low-resource languages, fostering responsible AI adoption across academic, industrial, and public sectors.

14. Bibliography

[1] Linyao Yang, Hongyang Chen, Senior Member, IEEE, Zhao Li, Xiao Ding and Xindong Wu Fellow, IEEE "Give us the Facts: Enhancing Large Language Models With Knowledge Graphs for Fact-Aware Language Modeling" , IEEE Tran, VOL.36,NO.7, pp 3091-3108, July 2024, DOI:10.1109/TKDE.2024.3360454

[2] Gilad Fuchs, Oded Zinman, Ido Ben-Shaul, "Practical Fact Checking System for LLMs" ,ACM on Web Conference 2025, Pages 2713 – 2716, May 2025, DOI: https://doi.org/10.1145/3701716.3717849

[3] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, Ting Liu, "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions", ACM Transactions, Volume 43, Issue 2, Pages 1 – 55, January 2025, DOI: https://doi.org/10.1145/3703155

[4] Rishabh Upadhyay, Marco Viviani, "Enhancing Health Information Retrieval with RAG by prioritizing topical relevance and factual accuracy", springer Discover Computing, Volume 28, April 2025, DOI: https://doi.org/10.48550/arXiv.2502.04666

[5] Katerina Hanikova, David Chudan, Vojtech Svatek, Peter Vajdecka, Raphaël Troncy, Filip Vencovsky, Jana Syrovátkova, "Towards Fact-check Summarization Leveraging on Argumentation Elements Tied to Entity Graphs" Pages 1473 – 1481, May 2024 DOI: https://doi.org/10.1145/3589335.3651914

[6] Junnan Liu, Wenlong Du, Qingquan Li, Xuewei Wang, Zhongjun Zhou, Jin Liu "S-Evaluator: Enhance Factual Consistency Evaluator with Adversarial Data Synthesized by Large Language Model", IEEE Xplore, pp:10116-10120, April 2024, DOI: 10.1109/ICASSP48485.2024.10448276

[7] Muneeswaran I, Advaith Shankar, Varun V, Saisubramaniam Gopalakrishnan, Vishal Vaddina, "Mitigating Factual Inconsistency and Hallucination in Large Language Models", ACM ISBN979-8-4007-0371-3/24/03, Pages 1169 – 1170, March 2024 DOI: https://doi.org/10.1145/3616855.3635744

[8] Haojie Zhuang, Wei Zhang, Weitong Chen, Jian Yang, Quan Z. Sheng, "Improving Faithfulness and Factuality with Contrastive Learning in Explainable Recommendation", ACM Trans., Volume 16, Issue 1, Pages 1 – 23, 2024 DOI: https://doi.org/10.1145/3653984

[9] Zheheng Luo a, Qianqian Xie b, Sophia Ananiadou , "Factual consistency evaluation of summarization in the Era of large language models" ScienceDirect Expert Systems With Applications, 2024 DOI: https://doi.org/10.1016/j.eswa.2024.124456

[10] Yunqi Xu, Tianchi Cai, Jiyan Jiang, Xierui Song, "Face4Rag: Factual Consistency Evaluation for Retrieval Augmented Generation in Chinese", ACM ISBN979-8-4007-0490-1/24/08 , Pages 6083 – 6094, August 2024, DOI: https://doi.org/10.1145/3637528.3671656

[11] Tianchi Cai, Zhiwen Tan, Xierui Song, Tao Sun, Jiyan Jiang, Yunqi Xu, Yinger Zhang, Jinjie Gu, "FoRAG: Factuality-optimized Retrieval Augmented Generation for Web-enhanced Long-form Question Answering", ACM ISBN979-8-4007-0490-1/24/08 Pages 199 – 210, August 2024, DOI: https://doi.org/10.1145/3637528.3672065

[12] Md Moinul Islam, Usman Muhammad, Mourad Oussalah, "Evaluating Text Summarization Techniques and Factual Consistency with Language Models", IEEE Xplore, December 2024, DOI: 10.1109/BigData62323.2024.10826032

[13] Arghya Kundu, Uyen Trang Nguyen, "Automated Fact Checking Using A Knowledge Graph-based Model", IEEE Xplore, March 2024, DOI: 10.1109/ICAIIC60209.2024.10463196

[14] Jiaju Kang, Weichao Pan, Tian Zhang, Ziming Wang, Shuqin Yang, Zhiqin Wang, "Correcting Factuality Hallucination in Complaint Large Language Model via Entity-Augmented", IEEE Xplore, September 2024, DOI: 10.1109/IJCNN60899.2024.10650208

[15] Xinyu Li, Yongbing Gao, Weihao Li, Lidong Yang, " A Text Generation Hallucination Detection Frame-work Based on Fact and Semantic Consistency" IEEE Xplore, March 2024, DOI:10.1109/AINIT61980.2024.10581604

[16] Hongjin Qian, Zhicheng Dou, Jiejun Tan, Haonan Chen, Haoqi Gu, Ruofei Lai, Xinyu Zhang, Zhao Cao, Ji-Rong Wen, "Optimizing Factual Accuracy in Text Generation through Dynamic Knowledge Selection" , arXiv, Aug 2023, DOI: https://doi.org/10.48550/arXiv.2308.15711

[17] Craig Thomson, Ehud Reiter, Barkavi Sundararajan, "Evaluating factual accuracy in complexdata-to-text", ScienceDirect Computer Speech & Language, Volume 80, May 2023, DOI: https://doi.org/10.1016/j.csl.2023.101482

[18] Sewon Min, Kalpesh Krishna, Xinxi Lyu, Mike Lewis, Wen-tau Yih, Pang Wei Koh, Mohit Iyyer, Luke Zettlemoyer, Hannaneh Hajishirzi, "FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation", arXiv, Oct 2023 DOI: https://doi.org/10.48550/arXiv.2305.14251

[19] Derek Tam, Anisha Mascarenhas, Shiyue Zhang, Sarah Kwan, Mohit Bansal, Colin Raffel, "Evaluating the Factual Consistency of Large Language Models Through News

Summarization", ACL, pages 5220–5255, July 2023, DOI: 10.18653/v1/2023.findings-acl.322

[20] Congqing He, Tien-Ping Tan, Sheng Xue, Yanyu Tan, "Explaining legal judgments: A multitask learning framework for enhancing factual consistency in rationale generation", ScienceDirect, Volume 35, Issue 10, December 2023, DOI: https://doi.org/10.1016/j.jksuci.2023.101868

[21] Yiyang Li, Xueyi Hao, Lei Li, "A Survey of Factual Consistency in Summarization from 2021 to 2023", IEEE Xplore, pp-84-89, March 2024, DOI: 10.1109/CAIT59945.2023.10468958

[22] Alejandro Martín , Javier Huertas-Tato, Alvaro Huertas-García, Guillermo Villar-Rodríguez, David Camacho, "FacTeR-Check: Semi-automated fact-checking through semantic similarity and natural language inference" , ScienceDirect Knowledge-Based Systems, pp-1-19, 2022, DOI; https://doi.org/10.1016/j.knosys.2022.109265

[23] Nayeon Lee, Wei Ping, Peng Xu, Mostofa Patwary, Pascale Fung, Mohammad Shoeybi, Bryan Catanzaro, "Factuality Enhanced Language Models for Open-Ended Text Generation", arXiv, 2022, DOI: https://doi.org/10.48550/arXiv.2206.04624

[24] Wenya Li, Xiaofei Zhou, Xu Bai, Shirui Pan, " Improving Factual Consistency of Dialogue Summarization with Fact-Augmentation Mechanism", IEEE Xplore, September 2022, DOI: 10.1109/IJCNN55064.2022.9891885

[25] Amanuel Alambo, Tanvi Banerjee, Krishnaprasad Thirunarayan, Mia Cajita, "Improving the Factual Accuracy of Abstractive Clinical Text Summarization using Multi-Objective Optimization", IEEE Xplore, 2022. DOI: 10.1109/EMBC48229.2022.9871798

[26] Jeongwan Shin, Seong-bae Park, Hyun-je Song," Token-level Fact Correction In Abstractive Summarization", IEEE Access, December 2022, DOI: 10.1109/ACCESS.2022.3233854

[27] Ruoyu Chen, Yan Li, Yuru Jiang , Bochen Sun, Jingqi Wang and Zhen Li "Fact-Aware Generative Text Summarization with Dependency Graphs", 2024, DOI: https://doi.org/10.3390/electronics13163230

[28] Ben Goodrich, Vinay Rao, Peter J. Liu, Mohammad Saleh, "Assessing The Factual Accuracy of Generated Text", CM SIGKDD Conference on Knowledge Discovery and Data Mining , 2019, DOI: https://doi.org/10.1145/3292500.3330955

[29] Yingqi Zhu, Yiyang Li, Lei Li, Dingxin Hu, Xueyi Hao, Dongsheng Chen, Xingyue Zhang, Zhejun Zhang, Yanquan Zhou, Marina Litvak, Natalia Vanetik, "Improving Factual Error Correction for Abstractive Summarization via Data Distillation and Conditional-Generation Cloze," IEEE Transactions on Audio, Speech and Language Processing, vol. 33, May 2025, DOI: https://doi.org/10.1109/TASLPRO.2025.3567744

[30] Anup Bera, Aditya Goel, "Factual Accuracy Checking by Validating Numerical Data in Financial Summaries," IEEE, 2024, DOI: https://doi.org/10.1109/BigData62323.2024.10825697

[31] Huan Rong, Gongchi Chen, Tinghuai Ma, Victor S. Sheng, Elisa Bertino, "FuFaction: Fuzzy Factual Inconsistency Correction on Crowdsourced Documents With Hybrid-Mask at the Hidden-State Level," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 1, pp. 167–180, Jan. 2024, DOI: https://doi.org/10.1109/TKDE.2023.3286097

[32] Yaping Chai, Haoran Xie, Joe S. Qin "Text Data Augmentation for Large Language Models: A Comprehensive Survey of Methods, Challenges, and Opportunities" , 2025, DOI: https://doi.org/10.48550/arXiv.2501.18845

[33] Vidula Magdum, Omkar Jayant Dhekane, Sharayu Sandeep Hiwarkhedkar, Saloni Sunil Mittal, Raviraj Joshi, "mahaNLP: A Marathi Natural Language Processing Library" , in IJCNLP 2023, DOI: 10.18653/v1/2023.ijcnlp-demo.5

[34] Raviraj Joshi, "L3Cube-MahaCorpus and MahaBERT: Marathi Monolingual Corpus, Marathi BERT Language Models, and Resources", in WILDRE June, 2022, DOI: https://aclanthology.org/2022.wildre-1.17/

[35] Qianren Mao, Jianxin Li, Hao Peng, Shizhu He, Lihong Wang, Philip S. Yu, Zheng Wang, "Fact-Driven Abstractive Summarization by Utilizing Multi-Granular Multi-Relational Knowledge," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 30, pp. 1665–1680, May 2022, DOI: https://doi.org/10.1109/TASLP.2022.3161157

[36] V. P. Kadam, S. A. Alazani, and C. N. Mahender, "A Text Summarization System for Marathi Language," International Journal of Computer Science and Information Technology, vol. 17, no. 9, pp. 566–581, Sep. 2022, DOI: 10.5281/zenodo.7073510

[37] Dhiraj Amin, Sharvari Govilkar, Sagar Kulkarni, Yash Shashikant Lalit, Arshi Ajaz Khwaja, Daries Xavier, Sahil Girijashankar Gupta, "Marathi-English Code-mixed Text Generation", arXiv preprint, 2023, DOI: https://doi.org/10.48550/arXiv.2309.16202

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Minakshi Ghorpade
Assignment title:	Assignment1
Submission title:	Designing a Fact-Aware Text Generation Framework for Marat...
File name:	ration_Framework_for_Marathi_Language_Topic_approval_Rep...
File size:	82.56K
Page count:	27
Word count:	6,498
Character count:	43,985
Submission date:	16-Aug-2025 03:09PM (UTC+0530)
Submission ID:	2730298550

ABSTRACT

The rapid advancement of Natural Language Processing (NLP) and Large Language Models (LLMs) has enabled the automatic generation of human-like text across diverse domains. However, these models often struggle with factual accuracy, frequently producing hallucinations, fabricated or unverifiable information. This is a critical issue, and especially for low-resource languages like Marathi due to the limited availability of large ready-to-use fact-annotated datasets, sophisticated linguistic tools, and established fact-verification systems present in other high-resource languages. Additionally, high-stakes sectors such as governance, education, journalism, and law are particularly vulnerable to inaccuracies, which can result in misinformation, and damage credibility and public trust.

This research aims to develop a Fact-Aware Text Generation Framework specifically for the Marathi language, designed to produce correct, trustworthy, reliable and contextually relevant content. The framework will be built by systematically analysing factual statements from verified Marathi sources, using contextual cues to construct the first comprehensive fact-annotated dataset in Marathi. This dataset will serve as the foundation for training and evaluating the model, ensuring that generated outputs are grounded in verified facts and significantly reducing the risk of misinformation.

The proposed framework aims to ensure and evaluate factual consistency, incorporating mechanisms to detect and reduce hallucinations by verifying generated text against a fact-annotated dataset. Automated metrics will assess accuracy, clarity, and completeness, along with domain-specific evaluations to ensure that it is usable for governance, health care, and education. The framework generated will lead to usable and trustworthy domain-adaptable Artificial Intelligent systems for Marathi, while serving as a scalable model for other lower-resourced languages, and bridging the gap between linguistic fluency and factual accuracy in AI-generated text.

i

Copyright 2025 Turnitin. All rights reserved.

Minakshi Ghorpade

Designing a Fact-Aware Text Generation Framework for Marathi Language

Assignment1

MultiCon25

Thakur College of Engineering and Technology (TCET)

Document Details

Submission ID
trn:oid:::1:3314517638

Submission Date
Aug 16, 2025, 3:08 PM GMT+5:30

Download Date
Aug 16, 2025, 3:21 PM GMT+5:30

File Name
ration_Framework_for_Marathi_Language_Topic_approval_Report.docx

File Size
82.6 KB

27 Pages

6,498 Words

43,985 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer
Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI generated as AI generated and AI paraphrased or likely AI generated and AI paraphrased writing as only AI generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

Designing a Fact-Aware Text Generation Framework for Marathi Language

ORIGINALITY REPORT

4%	4%	2%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	huggingface.co Internet Source	1%
2	aclanthology.org Internet Source	<1%
3	www.artpark.in Internet Source	<1%
4	Varshney, Neeraj. "Addressing Efficiency and Reliability Challenges in Natural Language Processing", Arizona State University, 2024 Publication	<1%
5	list.elra.info Internet Source	<1%
6	www.coursehero.com Internet Source	<1%
7	Nighojkar, Animesh. "An Inference-Centric Approach to Natural Language Processing and Cognitive Modeling", University of South Florida, 2024 Publication	<1%
8	Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong et al. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions", ACM Transactions on Information Systems, 2025 Publication	<1%
9	www.frontiersin.org Internet Source	

		<1 %
10	"Advances in Information Retrieval", Springer Science and Business Media LLC, 2025 Publication	<1 %
11	amatriain.net Internet Source	<1 %
12	sciencecast.org Internet Source	<1 %
13	trialsjournal.biomedcentral.com Internet Source	<1 %
14	Huan Rong, Gongchi Chen, Tinghuai Ma, Victor S. Sheng, Elisa Bertino. "FuFaction: Fuzzy Factual Inconsistency Correction on Crowdsourced Documents With Hybrid-Mask at the Hidden-State Level", IEEE Transactions on Knowledge and Data Engineering, 2024 Publication	<1 %
15	Liming Shao, Hong Yu, Wei Huang, Huiyuan Zhao, Lixin Zhang, Jing Song. "DeepSeek-Based Multi-dimensional Augmentation of Short and Highly Domain-Specific Textual Inquires for Aquaculture Question-Answering Framework", Aquaculture International, 2025 Publication	<1 %
16	Swati Swati, Dunja Mladenić, Marko Grobelnik. "An Inferential Commonsense-Driven Framework for Predicting Political Bias in News Headlines", IEEE Access, 2023 Publication	<1 %
17	arxiv.org Internet Source	<1 %
18	export.arxiv.org Internet Source	<1 %

19	itt-pgups.ru Internet Source	<1%
20	www.napier.ac.uk Internet Source	<1%

Exclude quotes	On	Exclude matches	< 8 words
Exclude bibliography	On		

University of Mumbai

Mahatma Jotirao Phule Bhavan, Vidyanagri, Santacruz (E), Mumbai - 400 098.

RECEIPT NO.: 3583 **DATE :** 22/08/2025

GENERAL FUND - I

Received from GHORPADE MINAKSHI the sum of Rupees

Paise Only

by cash being the amount of

TOPIC APPROVAL FORM FEE PHD

TOPIC APPROVAL FEES Rs. **********200.00

Total: Rs. **********200.00

Cash Receiver's Signature

Finance & Accounts Officer

1. Receipt subject to the realisation of Cheque / Draft etc., it tendered.
2. No. duplicate receipt will be issued.
3. The statement of marks should be collected from Ground Floor, Window No. 2 after 15 working days on production of the receipt and if not collected within 3 months no responsibility lies on the University.